THE EFFECT OF EXTREME WAGES ON AVERAGE WAGE VALUES

Lubos Marek

Abstract

As a basic indicator of the level of wages, arithmetic average is commonly used. It is an open secret that the average is far from an ideal measure, as it is strongly influenced by the outliers. In Czech Republic, we mean specially the wages over one hundred thousand Czech crowns per month cause the problem. We use several simulations and show how the average is changed when excluding the extremely high wages. We will review the use of basic quantiles as an appropriate measure of the wage levels. We will use median, lower and upper quantiles, 10% and 90% percentiles. We will use data for Czech Republic as a whole and then perform various comparisons. The calculations are performed over 1995-2016 years. Therefore, we are able to evaluate the trend of all computed statistics over time. We show that the contribution of extreme wages is small in terms of numbers, but the sum of these wages is great. The impact of these wages to the average wage is very significant.

Key words: average wage, extreme average, quantile measures, median

JEL Code: C530, F470

Introduction

It is generally known that arithmetic average values are strongly affected by outliers. Extremely low/high values substantially distort the averages. This observation is valid for any data, and the average wages in the Czech Republic are also affected. In this paper we show the effect of extremely high wages on average wage values, which are most frequently referred to as the characteristics of wages. Our data set is sufficiently large and structured, which enables us to recalculate the averages after excluding high wages, establishing the scope of the distortion. We will also show that the median values will remain practically unchanged even after the exclusion of high wages. Our data covers the period from 1995 to 2016; that is why we will also be interested in the time evolution of all considered and calculated characteristics.

1 Methodology

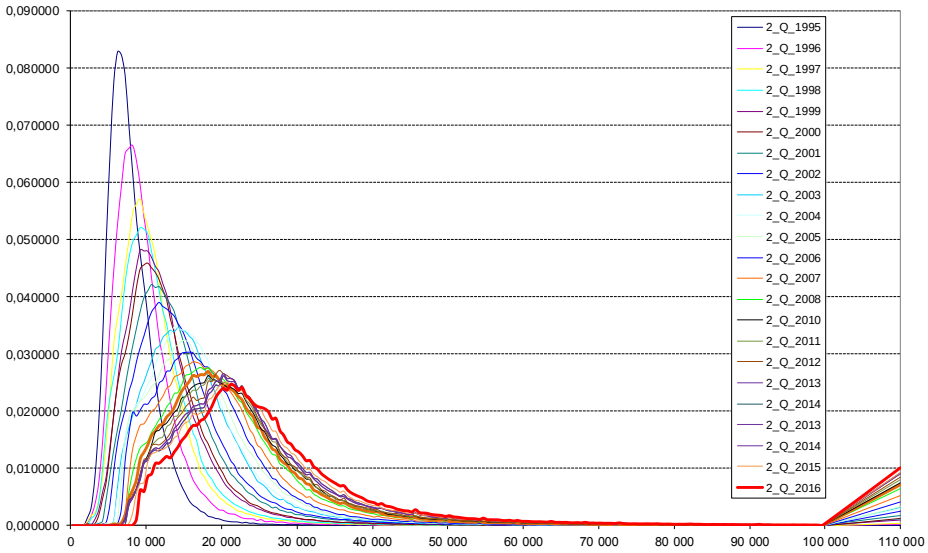
Our data is in the format of an interval histogram of frequencies. The interval resolution is 500 CZK, which is sufficiently fine for good-quality calculations directly from this table. The number of observations is large enough as well – from 321 thousand in 1995 to 2,119 thousand in 2016. These are average wages in the 2nd quarter of each year, which explains minor differences from the officially published average wages – the latter refer to each entire year. The data was collected by the company Trexima (www.trexima.cz). Our calculations require nothing more than the basic notions of descriptive statistics: weighted averages, group averages, and quantiles – cf. (Cyhelský 1981). The data has not been cleaned of the inflation, but this step is unnecessary for our comparisons. The conclusions would not be affected by such cleaning (we consider quotients).

2 Data Analysis – Wages in the Czech Republic

2.1 Proportion of High Wages

In Figure 1 we can see the polygon of frequencies for the average wages in the Czech Republic from 1995 to 2016. The right-hand side, showing high wages, is of particular interest. The proportion of high wages was negligible in the early years, but this proportion has been growing steadily. There was only 0.0068% (practically zero) of wages above 100,000 CZK ("high wages" below), but this proportion grew to 1.0125% in 2016.

Fig. 1: Polygon of wages



Source: author's own calculations

In Table 1 we show the proportion of high wages in the total number of wages, and also the proportion in the total amount. Figure 1 directly indicates that the influence of wages between 60,000 and 100,000 CZK is very small, regarding both quantity and amounts. The average will be affected mainly by wages above 100,000 CZK. It would be interesting to model the wages with the aid of a probability distribution. Such methods are studied by authors of (Bartosová and Longford 2014, Malá 2015, Marek and Vrabec 2016, Vrabec and Marek 2016). The question of wages is directly related to the income of households (Malá 2015 and 2016).

Tab 1: Proportions of high wages in total numbers and volumes of wages

	above 100,000	
year	% quantity	% amount
1995	0.007%	0.128%
1996	0.026%	0.333%
1997	0.045%	0.532%
1998	0.121%	1.386%
1999	0.090%	1.018%
2000	0.132%	1.557%
2001	0.168%	1.766%
2002	0.257%	2.648%
2003	0.326%	3.003%
2004	0.288%	2.722%
2005	0.337%	3.044%
2006	0.420%	3.795%
2007	0.532%	3.555%
2008	0.648%	4.142%
2009	0.717%	5.298%
2010	0.753%	5.342%
2011	0.807%	5.800%
2012	0.850%	5.803%
2013	0.916%	6.053%
2014	0.909%	5.736%
2015	0.929%	5.683%
2016	1.013%	5.871%

Source: author's own calculations

It is clear that the influence of high wages has been growing. This influence was negligible in the early years, but then it grew larger; we can see in 2016 values substantially larger. Even though the wages above 100,000 CZK made up just 1.013% of quantity in 2016, their proportion in the total amount was 5.871%. In other words, the proportion in the amount is more than five times the proportion in the quantity

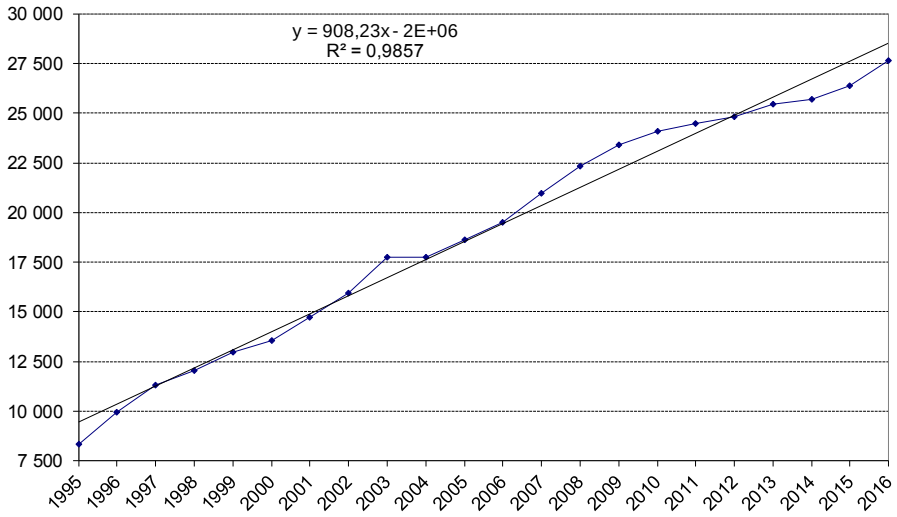
2.2 Average Wages

Let us first view the evolution of average wages shown in Figure 2.

It is clear from this Figure that the average value of the wages in the Czech Republic has been growing. The values fluctuate around the trend, but the evolution can be very well described by a trend line, with the index of determination equal to 0.9857. A more detailed analysis of the wage evolution can be found in (Marek 2010 and Marek 2013)

Let us now study the influence of the high wages on the average ones. We have at our disposal tables of wages; hence we are able to recalculate the averages after exclusion of wages above 100,000 CZK, or above any other limit we choose.

Fig. 2: Average wages in the Czech Republic



Source: author's own calculations

The influence of high wages on the average was negligible in the early years. In 1995, the average value went down by a mere 10 CZK after the exclusion of the wages above 100,000 CZK. In other words, the average was not significantly affected by the high wages and could be viewed as a good characteristic of wages.

However, the influence of the high wages on the average has been steadily growing. In the last year of our observations, the average went down by a substantial amount of 1,358 CZK after the exclusion of the wages above 100,000 CZK. In this instance the average becomes significantly affected by the high wages, and is no longer a good characteristic of wages.

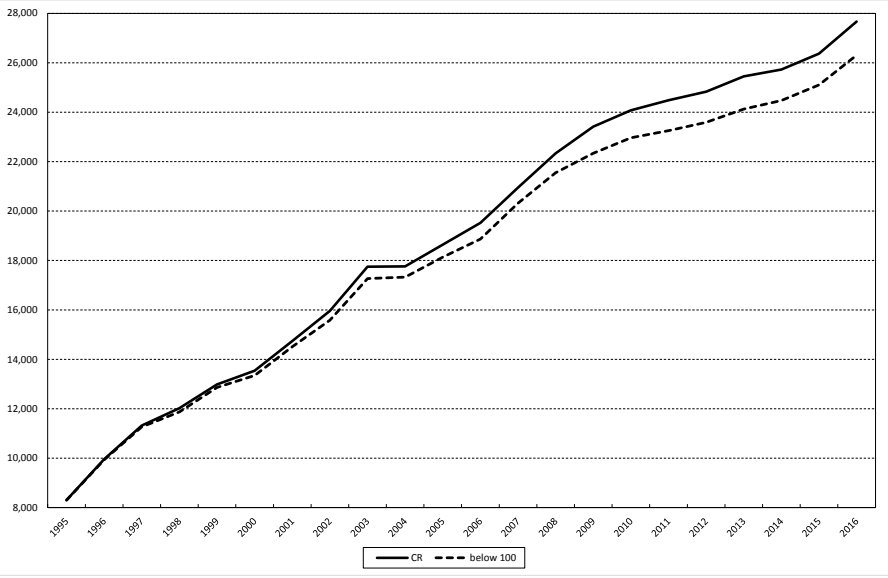
Tab. 2: Average wages after exclusion of high wages

year	CR	up to 100,000	difference
1995	8,311	8,301	10
1996	9,962	9,932	31
1997	11,322	11,266	55
1998	12,026	11,874	152
1999	12,982	12,862	121
2000	13,541	13,347	193
2001	14,743	14,507	236
2002	15,964	15,581	383
2003	17,748	17,271	477
2004	17,759	17,325	434
2005	18,640	18,134	506
2006	19,526	18,864	662
2007	20,953	20,316	637
2008	22,338	21,552	786
2009	23,418	22,338	1,080
2010	24,077	22,964	1,113
2011	24,484	23,251	1,232
2012	24,829	23,589	1,240
2013	25,448	24,129	1,319
2014	25,728	24,475	1,253
2015	26,369	25,104	1,265
2016	27,668	26,310	1,358

Source: author's own calculations

Fig. 3: Average wages in the Czech Republic

In Figure 3 we can see the time evolution of the differences between the official average wages (solid line) and those after the exclusion of the wages above 100,000 CZK (dashed line).



Source: author's own calculations

2.3 Quantile characteristics of wages

A question naturally arises: what characteristic would be suitable instead of the average? Quantile characteristics, and above all median, naturally come to mind in this context. Quantiles should not be affected by outliers, represented by high wages in our instance. Table 3 confirms this approach. The reason is simple: there are few high wages, and their exclusion will therefore not significantly affect the complete set of the observed values. In the early years, the median values do not undergo any changes at all. With the growing proportion of the high wages, even the median values are slightly changed by their exclusion; but this effect is only present in the last 10 years. We can recommend without hesitation that the median value should be published together with the average value.

Tab. 3: Median after exclusion of high wages

year	median CR	median 100	difference
1995	7,500	7,499	1
1996	8,956	8,957	-1
1997	10,171	10,175	-4
1998	10,563	10,556	7
1999	11,506	11,500	5
2000	11,860	11,852	8
2001	12,901	12,892	8
2002	13,857	13,841	15
2003	15,519	15,494	25
2004	15,789	15,771	18
2005	16,432	16,406	26
2006	17,143	17,108	35
2007	18,185	18,341	-156
2008	19,267	19,437	-170
2009	20,138	20,287	-149
2010	20,753	20,841	-88
2011	21,020	20,942	78
2012	21,319	21,237	82
2013	21,779	21,687	92
2014	22,074	21,979	95
2015	22,658	22,567	91
2016	23,757	23,649	108

Source: author's own calculations

The columns have the following meanings:

"median CR" – the officially published median values for all wages;

"median 100" – the median value for the wages up to 100,000 CZK; and

"difference" – "median CR" minus "median 100".

Conclusions

The presented calculations unambiguously indicate that the average is not a proper characteristic of wages. It is significantly affected by high wages, especially in the recent years. In the beginning years of our observations, there were very few high wages, which did not affect the average value; or rather, their effect was negligible. The influence of the high wages has, however, been growing and in the recent years they affect the average value of wages quite substantially. Their proportion is very small regarding the quantity, but the proportion in the total amount is much higher. The median appears to be a suitable replacement for the average – it is not overly sensitive to high wages because only 1% of wages are above 100,000 CZK. Hence, we recommend accompanying the references to averages by references to median values.

Acknowledgments

This paper was written with support from the Czech Science Foundation, Project No. P402/12/G097 "DYME – Dynamic Models in Economics."

References

[1] Bartosová, J. and Longford, N. T. (2014) "*A study of income stability in the Czech Republic by finite mixtures'*, Prague Economic Papers, vol. 23, no. 3, pp. 330–348. http://dx.doi.org/

[2] Cyhelský, L. (1981). Úvod do teorie statistiky (In Czech; Introduction to Theory of Statistics). Praha. SNTL.

[3] Malá, I. (2015) "*Vícerozměrný pravděpodobnostní model rozdělení příjmů českých domácností" (In Czech; Multidimensional probabilistic model for Czech household income distribution)*, Politická ekonomie, vol. 63, no. 7, pp. 895–908. http://dx.doi.org/10.18267/j.polek.1040.

[4] Malá, I. (2015). "*Equivalence scales for evaluation of equivalised income of Czech households". In: The 9th International Days of Statistics and Economics (MSED 2015) [online]*. Praha, 10.09.2015 – 12.09.2015. Slaný : Melandrium, 2015, s. 1069–1078. ISBN 978-80-87990-06-3. Available from: http://msed.vse.cz/msed_2015/article/171-Mala-Ivana-paper.pdf.

[5] Malá, I. (2016). *Vliv definice spotřebních jednotek na ekvivalizované příjmy domácností v České republice*. Acta Oeconomica Pragensia [online]. 2016, roč. 24, č. 3, s. 53–67. ISSN 0572-3043. DOI: 10.18267/j.aop.536.

[6] Marek, L. and Vrabec, M. (2016). Using mixture density functions for modelling of wage distributions. *Central European Journal of Operations Research* [online]. 2016, roč. 24, č. 4, s. 389–405. ISSN 1435-246X. DOI: 10.1007/s10100-015-0409-4.

[7] Marek, L. (2010) *Analýza vývoje mezd v ČR v letech 1995-2008* (*In Czech; Analysis of wage evolution in the Czech Republic from 1995 to 2008*). Politická ekonomie, vol 58, no. 2, pp. 186–206.

[8] Marek, L. (2013) '*Some Aspects of Average Wage Evolution in the Czech Republic'*, In: *International Days of Statistics and Economics*, Prague, Slaný: Melandrium, pp. 947–958.

[9] Trexima, (2016) http://www.trexima.cz

[10] Vrabec, M., Marek, L. (2016) "*Model for distribution of wages*". In: *AMSE 2016* (*Applications of Mathematics and Statistics in Economics*) [online]. Banská Štiavnica, 31.08.2016 – 04.09.2016. Banská Bystrica : Občianske združenie Financ, 2016, s. 378–386. ISBN 978-80-89438-04-4. ISSN 2453-9902.

Contact

Lubos Marek
University of Economics, Prague
W. Churchill Sq. 4, 130 67 Prague 3
marek@vse.cz